Title: Development of neural network emulations of model physics components for improving the computational performance of the NCEP seasonal climate forecasts

Proposed period: 36 months, 03/01/06 – 02/28/09

Lead P.I.: Michael Fox-Rabinovitz, ESSIC/University of Maryland
P.I.: Vladimir Krasnopolsky, NOAA/NCEP/EMC/SAIC
Co-I.: Stephen Lord, NOAA/NCEP/EMC
Co-I.: Yu-Tai Hou, NOAA/NCEP/EMC

Abstract:
Consistently with the CTB goal "to accelerate the transition of research and development into improved operational climate forecasts, products and applications", the proposed collaborative research will lead to *improving the computational performance for operational seasonal climate predictions*. The proposed study will take an advantage of the NCEP CTB that "will provide an operational testing environment to support … applied research and development projects that will result in a direct influence on NOAA climate forecast operations, to be carried out jointly by scientists from operational centers and the broader research community."

This innovative research will be devoted to developing accurate and fast emulation of model physics using statistical learning techniques, namely neural networks (NN). It will be aimed at becoming an integral part of the operational NCEP CTB climate forecasting system. The proposed research will be based on the already existing and ongoing mutual collaboration between the NCEP and University of Maryland participants and their respective groups.

The proposed collaborative research is considered by the proposers as a strategic scientific and methodological study with immediate practical applications to the NCEP operational system.

Research highlights for the report period:

- The current NCEP CFS model (GFS/MOM3) with NCEP's versions of RRTM1 LWR and RRTM2 SWR (modified from AER Inc.'s RRTMG LW and SW radiation codes) have been used
- Creation of NN training data sets for both LWR and SWR has been developed and adjusted for the *coupled* NCEP CFS model
- The NN methodology and experimentation framework have been developed and refined
- Refined NN emulations of both LWR and SWR have been developed and their accuracy estimated against the original LWR and SWR
- A new measure of similarity for climate simulations and seasonal predictions, i.e., comparison with the interval model variability, has been introduced for comprehensive validation
- Refined NN emulations for full model radiation (i.e. using both LWR NN and SWR NN) have been validated including comparison against the interval model variability; it was done for the parallel control (using the original LWR and SWR) and NN runs for the 17-year (1990-2006) climate simulation and seasonal predictions

- The refined NN emulations for **full model radiation have been transferred/made available for the independent application by another EMC group to medium-range predictions with NCEP GFS**; the application has shown positive/promising results and now is being further tested in the pre-operational mode by the EMC implementation group.
- Journal and conference papers have been published, and the CTB seminars have been given.

The above developments allowed us to use **NN emulations for the entire model radiation block**, i.e., using LWR NN and SWR NN simultaneously for the NCEP CFS model (GFS/MOM3), which is **the major goal and accomplishment of the project final report**.

1. Background Information

The CTB project has been dedicated to a novel research on development of *accurate and computationally efficient neural network (NN) emulations of climate model physics*, aimed at application to the NCEP climate forecasting system (CFS). The research effort is based on the participants' solid methodological and practical expertise and experience on successful development of robust NN emulations for the most time consuming part of model physics, the long-wave and short- wave radiation (LWR and SWR).

The project research has been focused on a practical *transition* of the NN emulation framework to the NCEP CFS for operational seasonal climate predictions. The framework includes methodologies on development NN emulation and their validation, in terms of both the accuracy of NN approximation and seasonal predictions.

The project activities have been organized as specific *tasks* dedicated to *the transition of the developed NN approach into the NCEP CFS model*.

2. Description of the tasks for the report period

The tasks have been devoted to development NN emulations of the NCEP CFS model radiation, the most time consuming block of the model.

The major effort was dedicated to the following developments:
Reporting Period: 12/01/2009 - 02/28/2010**:** Development and refinement of NN emulations for the RRTM1 Long-Wave Radiation (LWR) and RRTM2 Short-Wave Radiation (SWR) parameterizations of the NCEP CFS model (GFS/MOM3). **Successful completion of of the project dedicated to transferring the NN full radiation emulations into the NCEP CFS.**

1. Completion of development and refinement of the NN methodology for emulating model LWR and SWR.
2. Completion of development and refinement of experimentation and validation framework.
3. Completion of development and refinement of creating representative data sets for NN training.
4. Completion of development and refining of NN emulations for the NCEP CFS (GFS/MOM3) model (T126L64) for full model radiation (i.e., for RRTM1 LWR and RRTM2 SWR parameterizations) and estimation of their accuracy against the original LWR and SWR.

5. Introduction of a new measure of similarity for climate simulations and seasonal predictions, i.e., comparison with the internal model variability, for comprehensive validation
6. Performing and validating 17-year (1990-2007) climate simulations and seasonal predictions (for seasons 1-4) using the NCEP CFS (GFS/MOM3) model (T126L64) with NN emulations of RRTM1 LWR and RRTM2 SWR, including comparison against the interval model variability and using the NCEP validation package (obtained from Dr. S. Suru).
7. Transferring/making available the refined NN emulations for full model radiation for the independent application by another EMC group to medium-range predictions with NCEP GFS

   All the above tasks have been successfully completed during the report period.

The new developments during the report period include:

Completion of development and <u>refinement of the NN methodology for emulating model radiation</u> (Krasnopolsky et al. 2008 a-e)
Decadal Climate Simulations Using Accurate and Fast Neural Network Emulation of Full, Long- and Short Wave, Radiation (Krasnopolsky et al. 2010, 2009, 2008d)
Compound Parameterization (CP) with Quality Control (QC) for larger errors (Krasnopolsky et al. 2008b)
NN Ensembles for improving the accuracy of NN emulations (Krasnopolsky et al. 2008e)

<u>**Refinement of experimentation and validation framework**</u> (Krasnopolsky et al. 2010)

The major new methodological and experimental refinement has been introduction of a new measure of similarity for climate simulations and seasonal predictions, i.e., comparison with the internal model variability, for comprehensive validation

<u>3. Creating representative data sets for training NN</u>

Creation of NN training data sets has been developed and adjusted for the *coupled* NCEP CFS model by defining the period and sampling of the NN training data needed for the *coupled* model. This was a practically important effort beneficial for increasing the overall accuracy of NN emulations. The training set was selected from the 17-year (1990 – 2006) run produced with the current NCEP CFS T126L64 model. Half of the data is used for training and another half for validation or NN accuracy estimation vs. the original LWR. Every 2 weeks, one day, i.e. eight 3 hourly global files have been recorded; totally – 2,080 files. 100 events/profiles have been selected randomly from each 3 hourly global file. Thus, totally – 208,000 events/profiles is used for NN training and the same number of independent events/profiles for NN validation.

<u>4. Completion of development and refinement of NN emulations for the RRTM1 LWR and RRTM2 SWR parameterizations and validation of their accuracy (Krasnopolsky et al. 2010)</u>

Summary

During the report period for this project, the approach to accurate and fast calculating model physics using neural network emulations has been further refined by the project collaborators for both long-wave and short-wave radiation parameterizations, or for the *full model radiation*, the most time-consuming component of model physics. The neural network emulations of full model radiation have been included into the NCEP *coupled* CFS (Climate Forecast System) with *high resolution* and *time dependent* CO_2. The high complexity of NCEP CFS required introducing further adjustments to the neural network emulation methodology. Validation of the approach for the NCEP CFS has been done through a decadal (17-year) climate simulation and seasonal predictions. The developed highly accurate neural network emulations of long-wave and short-wave radiation parameterizations are on average 16 and 60 times faster than the original/control long-wave and short-wave radiation parameterizations, respectively. A detailed comparison of parallel decadal climate simulations and seasonal predictions performed with the original NCEP model radiation parameterizations and with their neural network emulations is presented. Almost identical or close results for model prognostic and diagnostic fields have been obtained for the parallel decadal simulations and seasonal prediction that justifies the practical use of efficient neural network emulations of full model radiation for climate simulations and seasonal predictions.

Bulk Approximation Error Statistics

To ensure a high quality of representation of the LWR and SWR processes, the accuracy of their NN emulations has been carefully investigated. The NN emulations have been validated against the original NCEP CFS LWR and SWR parameterizations. To calculate the error statistics presented in Tables 1 and 2 and in Fig.1, the original parameterizations and their NN emulations have been applied to the validation data set. Two sets of the corresponding HR (heating rates) profiles have been generated for both LWR and SWR. Total and level bias (or a mean error), total and level RMSE, profile RMSE or PRMSE, and σ_{PRMSE} have been calculated.

Table 1 shows bulk validation statistics for the accuracy of approximation of heating rates (HR). It also shows the average computational performance or the speedup, which depends significantly on cloud situations as shown in Table 2, obtained for a single processer configuration for the best, in terms of both the accuracy and performance, developed NN emulations for the NCEP CFS LWR and SWR. For comparison, the information on the NN emulations for NCAR CAM LWR and SWR (Krasnopolsky et al. 2008a) is also presented in Table 1. Total statistics show the bias, RMSE, PRMSE, and σ_{PRMSE} for the entire 3-D HR fields. Also, layer statistics for the top and bottom atmospheric layers are included to illustrate the accuracy of NN emulations in the areas of the increased non-linearity. Although the two models as well as their embedded radiation parameterizations are different, comparisons between NCAR CAM (with 26 vertical layers) and NCEP CFS (with 64 vertical layers) allow us to observe a general dependence of the NN accuracy on the model vertical resolution (see also error profiles shown in Fig.1). In additional to the GCM versions of LWR and SWR (RRTMG), statistics of a full scale version of RRTM LWR (RRTMF) is also included in the table for references.